

Automated Driving

Getting and Keeping the Human in the Loop

Insights

- Automated driving is an interesting domain for HCI, as it involves a safety-critical physical system that is used by nonprofessional users.
- When humans use automated vehicles, they can experience mode confusion. Mode confusion can be studied systematically using hidden Markov models.
- We propose a framework for studying the negotiated attention interleaving between driving and non-driving tasks.

Automated vehicles are sometimes seen as fully replacing a human driver, particularly in the popular press and in nonscientific conversation. However, although there are successes in automated vehicles, no systems can yet achieve this. Instead, we argue that human-automation interaction in the context of vehicles should be considered as a partnership between the human and the automated system. That is, instead of either or, the system includes the human, and the human, when appropriate, also actively takes part in some of the tasks associated with driving.

Human-machine interaction has been a critical aspect of driving research for decades. After all, safe manual control of a vehicle requires well-designed interactions with the car, and with any devices that are built into or brought into the car [1]. Interestingly, with the advent of automated driving, human-computer interaction (HCI) and, more broadly, the role of the human in the vehicle, remains an important aspect of vehicle design.

We view HCI in automated driving in the wider context of human interactions with automated systems. We recently reviewed research on human-automation interaction over the past 50 years [2]. In that interval, the research within the field has been expanding. Automation is applied more and more in the following contexts:

- Time-sensitive, safety-critical scenarios, such as power plants, transportation, and medicine
- Use by nonprofessional users, for example in applications such as robot vacuums, personal medicine, and intelligent chatbots
- Embodied, situated systems. That is, in physical systems that interact with the world around them in a dynamic way, for example, in drones, robot vacuums, and airplanes.

Semi-automated vehicles involve all three of these characteristics. For this reason, these vehicles are an interesting specific domain to study human-automation interaction. At the same



time, it is also a domain in which a lot can be learned from other domains that share these characteristics [3].

Given the expansion of the field of human-automation interaction, there is a large potential for the field of human-computer interaction to contribute. For driving specifically, automation has the potential to open mobility options to a wider set of users. For example, people with physical limitations or accessibility issues (e.g., limited sight, or inability to use hands or legs for movement) might be able to travel in an automated vehicle. An issue is how to design these vehicles and systems such that they are indeed accessible to different user populations. Furthermore, automated vehicles hold the promise of allowing drivers to reclaim some of the time they currently devote to driving. Human-computer interaction must be carefully designed in order to allow users to engage in nondriving tasks while automation is in charge, so that they can be productive and/or can enjoy themselves, but also in

order to allow them to safely return to the driving task when this is necessary.

Another potential for HCI has to do with training. In the past, automated systems were mostly used by experienced users with extensive, dedicated training on how to use such systems. For example, part of the job of an airplane pilot or factory operator is to go through appropriate training. Driving on public roads also requires a driver's license, and therefore some basic knowledge and skills (and ideally, training). However, safe use of automated technology is not part of a driver's test. This makes it essential that the eventual use of such technology is intuitive and clear, which HCI insights can foster.

Moreover, although the hardware of any individual car might not change overnight, the software might, through updates. How can we keep drivers informed about how their car works, what they can expect of the car, and what the car expects of them? For example, do the humans just monitor the behavior

of the car, or do they need to take active control? And under what circumstances is this needed?

Overall, just like the field of human-automation interaction at large (see [2]), for automated driving there is a need to keep the human driver in the loop. Human-computer interaction is specialized in carefully studying and designing systems for human use. HCI methods, theories, and insights will be critical for the safe exploitation of automated vehicles.

DYNAMIC HUMANS AND DYNAMIC SYSTEMS: MODE CONFUSION

Another reason to consider the human explicitly in the design of automated systems, and to keep the human driver in the loop, is that humans, systems, and the environments in which they operate are dynamic. The combination of these dynamic systems can lead to so-called mode confusion. Let's unpack this further.

Humans are dynamic in multiple ways. First, humans might use systems differently than designers intended. For example, one motivation for automated driving is to minimize the potential for human error. Systems try to achieve this by reducing the number of basic vehicle tasks that the driver has to perform, to allow them to focus more on other important driving-related tasks, such as monitoring traffic and anticipating dangerous situations. Unfortunately, in practice, studies have observed that in situations where more tasks are transferred from the human to the machine, people perform other non-driving-related tasks more frequently. The result is reduced (instead of improved) situational awareness, and slower (instead of faster) responses to safety-critical events. This can be described as an “irony of automation” [4], where the introduction of automation does not solve an issue (e.g., reduce human errors), but instead changes human behavior and introduces new, different problems. It can also be seen in a wider context of people reappropriating technology that is familiar from other forms of technology use in HCI. The human remains dynamic.

A second way in which humans are dynamic is in their learning and unlearning of skills, habits, and knowledge. Together, these shape expectations of a system and specific interaction patterns. However, knowledge (of a system’s functioning) might be incomplete and habits might prevent the learning of new skills, and these together might limit appropriate system use. Active training might help to overcome these biases, but might not always be immediately available (for example: might not be there immediately after a software update).

Similar to humans, automated systems and the environments in which they operate are also dynamic. First, as an instantiation of artificial intelligence, automated systems often learn from their environment, which can shape and change their responses over time. Second, automated systems are typically developed for use in specific contexts. In the case of automated vehicles, the Society of Automotive Engineers (SAE) has identified specific levels of automation for the functionality of the car, which can be used in specific “operational design domains,” or contexts. However,

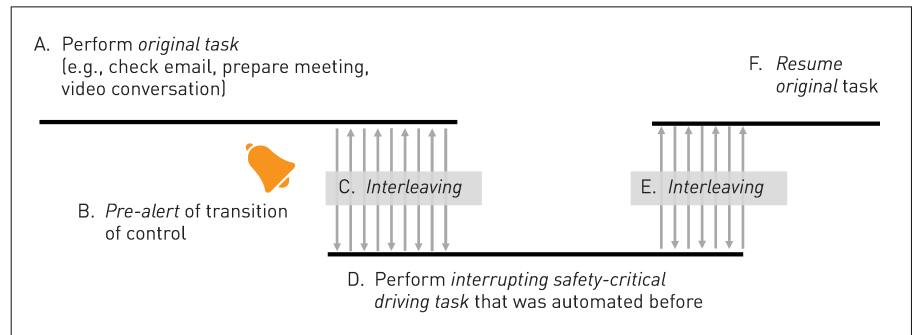


Figure 1. Handling a transition of control in automated driving from an interruption process. This figure captures the main processes. An even more refined framework with 10 stages is described in [6].

the context in which a car is driving can change over time and space, and thereby the system’s functioning and reliability can change over time and space. For example, adaptive cruise control might maintain the speed of a vehicle and a safe distance to cars in front of it on regular highways, under normal traffic and weather conditions. Yet what is safe might change with context: If there is suddenly heavy snowfall, the system might fail to act appropriately, and responsibilities that the car had (e.g., maintaining a safe distance to the car in front of it) might suddenly be transferred to the human.

The combination of dynamic humans that use dynamic systems can create situations in which there is confusion. In particular, humans might have assumptions about the mode that a system operates in: what functionality the system has, and what functionality it does not have. As described above in the adaptive cruise control example, the mode of operation might change abruptly. If the user is not alerted, or if they overlook an alert, the result is mode confusion: a discrepancy between the actual system state and the human’s belief about the state.

We recently argued that there is a need for formal tools to systematically analyze and design for the avoidance of mode confusion, and we presented a hidden Markov model framework as a candidate technique [5]. Hidden Markov models can explicitly separate things that are observable (e.g., system state) and latent inferences that are made (e.g., a human belief about the state). Moreover, one of the core characteristics of hidden Markov models is that states (and associated beliefs) can change, and that future states depend on the current state. This makes them an ideal candidate for describing mode confusion, as there

is the ability to explicitly distinguish human belief from system functionality and how these might change over time. Using such a framework can therefore force designers and analyzers to separate states from beliefs and to consider transitions explicitly.

NEGOTIATED ATTENTION INTERLEAVING

Given that humans and machines are dynamic, and that system states can transition, it is important to develop accurate frameworks, theories, and systems to negotiate and communicate transitions, and to guide attention in human-automation interaction settings [2]. One area that can provide inspiration is interruption research. We recently proposed a way in which frameworks from this area can be applied to automated-driving settings [6].

Interruption studies have been influential in human-computer interaction, and also in driving research. There are detailed theoretical models and frameworks, critical experimental tests, and extensive observational studies of how interruptions are handled in professional settings. However, in the typical interruption study, the assumption is that people are working on an urgent and sometimes safety-critical task (such as driving manually), which is occasionally interrupted by another task (such as a phone call).

This changes with an automated-driving setting. If automated vehicles live up to their promises, then their ability to drive with minimal human assistance over long stretches of road will increase, and the need for the human to contribute to the driving task will be reduced. Therefore, humans might start working on other, nondriving tasks, such as checking email, preparing for a meeting, or having a video conversation with a

remote conversant [7]. These tasks might then be experienced as very important to the human. Occasionally, though, the human will need to drive. However, when the car issues a request to the human to start driving (often referred to as a “transition of control”), the human might experience this as an interruption of the original nondriving task that they were working on. So, instead of driving being interrupted by other tasks (like a phone call), in automated vehicles, other tasks are interrupted by driving.

We argued that this interruption can then be described as a process, which is summarized in Figure 1. Based on interruption theory, this process can be decomposed into multiple stages: after an initial alert or signal (Figure 1B), the human might interleave their attention between finishing up their original (nondriving) task and orienting to the driving scenario (C) before taking full control of the car (D). After handling the safety-critical scenario (e.g., navigating the car through roadworks), they might also slowly hand back control to the car, which might again involve interleaving (E): While resuming an original task, the human driver might occasionally monitor whether the car has resumed the driving task correctly. Figure 1 shows this process at a high level, but even more detailed stages can be identified. In our full paper we identify 10 potential stages and discuss what is known about each of these stages from previous research [6].

Although we make a case to describe transitions of control as going through multiple stages, the field at large is mostly focused on only two stages: 1) the moment when the car alerts the human that it is requesting a transition of control, and 2) the moment when this transition actually happens through physical action. Intermediate stages are mostly described informally. Moreover, there is a focus on minimizing the time between the initial alert and the physical action. Although intuitively it might seem beneficial to minimize this interval, interruption literature suggests that it is beneficial to also consider more *careful* (and therefore, sometimes slower) interruption handling. Taking some time to finish the original task and to prepare for another interrupting task has multiple benefits. For example, it has the potential to reduce workload, stress, and mental distraction. Thereby, careful

preparation (which we argue can be done through interleaving of attention) can allow the human to be better prepared to handle a complicated traffic scenario.

Our distinction of multiple stages in the interruption process can also aid in the understanding of and design for transitions of control. Specifically, instead of simply designing an intervention to aid transitions of control, designers can think about which specific *stage* their intervention is targeting. This makes contributions more precise. Moreover, interruption research suggests that interventions at one stage of the interruption process can have consequences for all subsequent stages. Therefore, consideration of the full chain of events, and being explicit about all stages, is essential in order to make interventions effective. Our framework allows for a discussion of interventions in this wider context.

CONCLUSION

The field of human-automation interaction, and of human interaction with automated vehicles, is at an exciting stage. There are many technical developments, but there is also more and more consideration for the human experience. Moreover, humans remain crucial players in many automation situations, including automated driving. Given that both humans and automated systems are dynamic, there is potential for the full spectrum of HCI methods and theories to provide insight on a wide spectrum of users and scenarios.

HCI and its parent and partner disciplines can provide crucial insights about human behavior. For example, engineering developments can inform the design of automated systems, and theory on interruption handling can inform interventions for human-automation interaction. However, these other domains can also learn from HCI studies of human-automation interaction. For example, in our refinement of the framework of interruptions to fit the scenario of automated driving, we introduced the concept of interleaving, whereas traditional interruption theories often handle interruptions as (forced, often not deferrable) task switches. In that way, the original disciplines can also learn more about their theories from application to a domain. This makes for a fruitful and

interesting exchange of ideas between different disciplines, and we look forward to the many contributions that the field of HCI will make in this way to human-automation interaction.

ACKNOWLEDGMENTS

Andrew Kun was supported in part by NSF grant CMMI-1840085.

ENDNOTES

1. Kun, A.L. Human-machine interaction for vehicles: Review and outlook. *Foundations and Trends in Human-Computer Interaction* 11, 4 (2018), 201–293.
2. Janssen, C.P., Donker, S.F., Brumby, D.P., and Kun, A.L. History and future of human-automation interaction. *International Journal of Human-Computer Studies* 131 (2019), 99–107.
3. Meschtscherjakov, A., Tscheligi, M., Pfleging, B., Sadeghian Borojeni, S., Ju, W., Palanque, P., Riener, A., Mutlu, B., and Kun, A.L. Interacting with autonomous vehicles: Learning from other domains. *Extended Abstracts of the 2018 SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, 2018, W30.
4. Bainbridge, L. Ironies of automation. *Automatica* 19, 6 (1983), 775–779.
5. Janssen, C.P., Boyle, L.N., Kun, A.L., Ju, W., and Chuang, L.L. A hidden Markov framework to capture human-machine interaction in automated vehicles. *International Journal of Human-Computer Interaction* 35, 11 (2019), 947–955.
6. Janssen, C.P., Iqbal, S.T., Kun, A.L., and Donker, S.F. Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies* 130 (2019), 221–233.
7. Kun, A.L., Meulen, H. van der, and Janssen, C.P. Calling while driving using augmented reality: Blessing or curse? *Presence: Virtual and Augmented Reality* 27, 1 (2019), 1–14.

📌 **Christian P. Janssen** is an assistant professor of experimental psychology at Utrecht University. He received his Ph.D. in human-computer interaction from UCL (2012). His research is dedicated to understanding human attention, human-automation interaction, and adaptive behavior. He was one of the general chairs of the ACM SIGCHI's Auto-UI 2019 conference.
→ c.p.janssen@uu.nl

📌 **Andrew L. Kun** is professor of electrical and computer engineering at the University of New Hampshire. He received his Ph.D. in electrical engineering from the University of New Hampshire (1997). His research focuses on human-computer interaction in vehicles. He is the interim ACM SIGCHI vice president for conferences.
→ andrew.kun@unh.edu